Vocabulary Development & Governance Scenarios

caBIG Vocabulary/CDE

Work Group

7/19/2004

The Issue

- caBIG will need terminology resources
- We can standardize on existing resources
    - LOINC, SNOMED/CT, MedDRA, MGED Ontology, NCI Thesaurus, VA/NDF, etc.
    - Some Governance questions are how do we:
        - Select them ?
        - Deploy them ?
        - Influence their future development ?
        - Extend them ?

Selection

- What is the right balance in caBIG between
  - Adherence to emerging standards suites
    - HL7 v3, CHI
  - Compatibility with legacy systems
    - MIS, HIS, LIS, Rad

Deployment

- Should caBIG assume terminology servers:
  - At each Center, serving that Center's caBIG needs?
  - Shared caBIG vocabulary services available to Centers via a GRID or other distributed architecture?
  - A mix of terminology embedded in Center systems (no terminology service)?
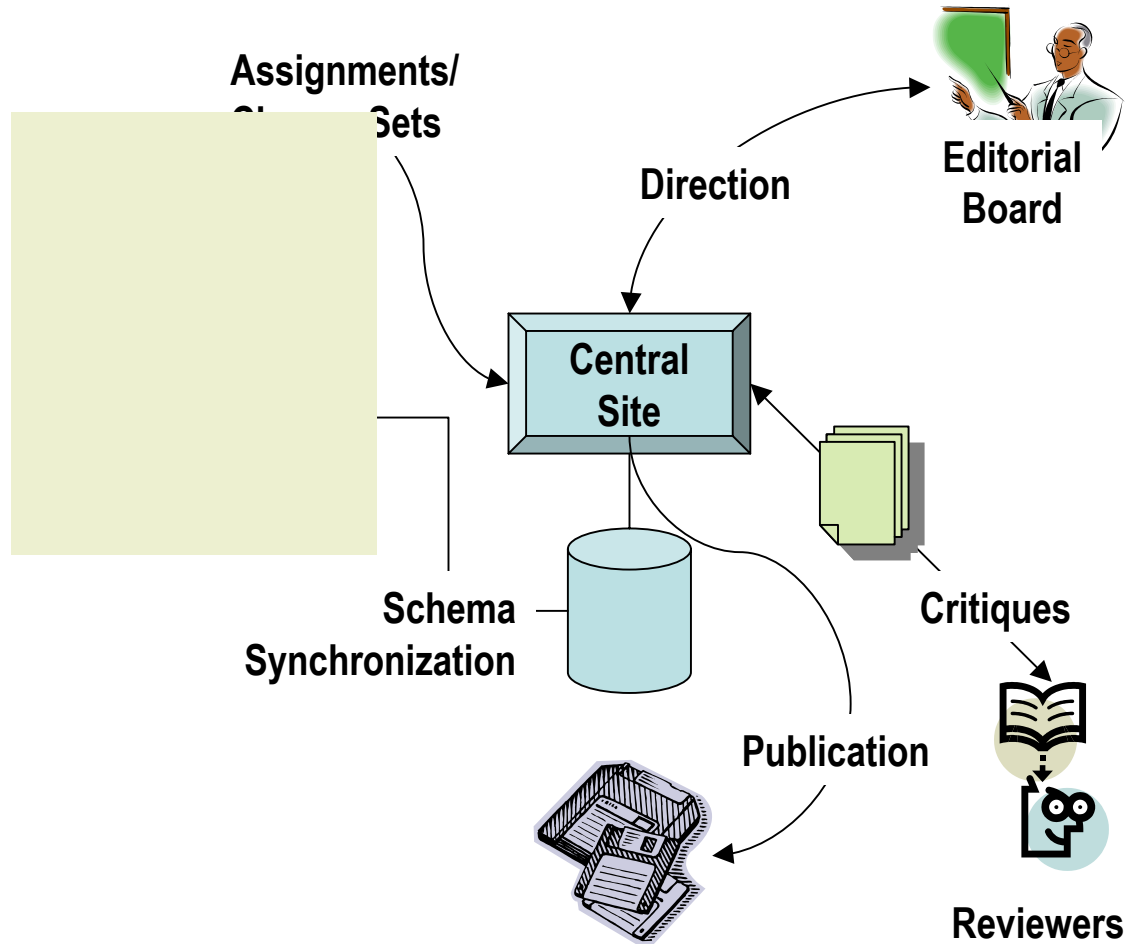
Future Development

- How should caBIG influence CAP, Regenstrief, NCI, VA and other producers of standard terminologies?
  - Work thru Center and NCICB staff who are already working with these groups?
  - Is tighter integration of caBIG needs and NCI EVS development desirable?

Vocabulary Development

- Extension of standard terminologies or creation of specialized terminology
- What is the right balance of control and freedom in development of extensions, new terminology for use by caBIG?
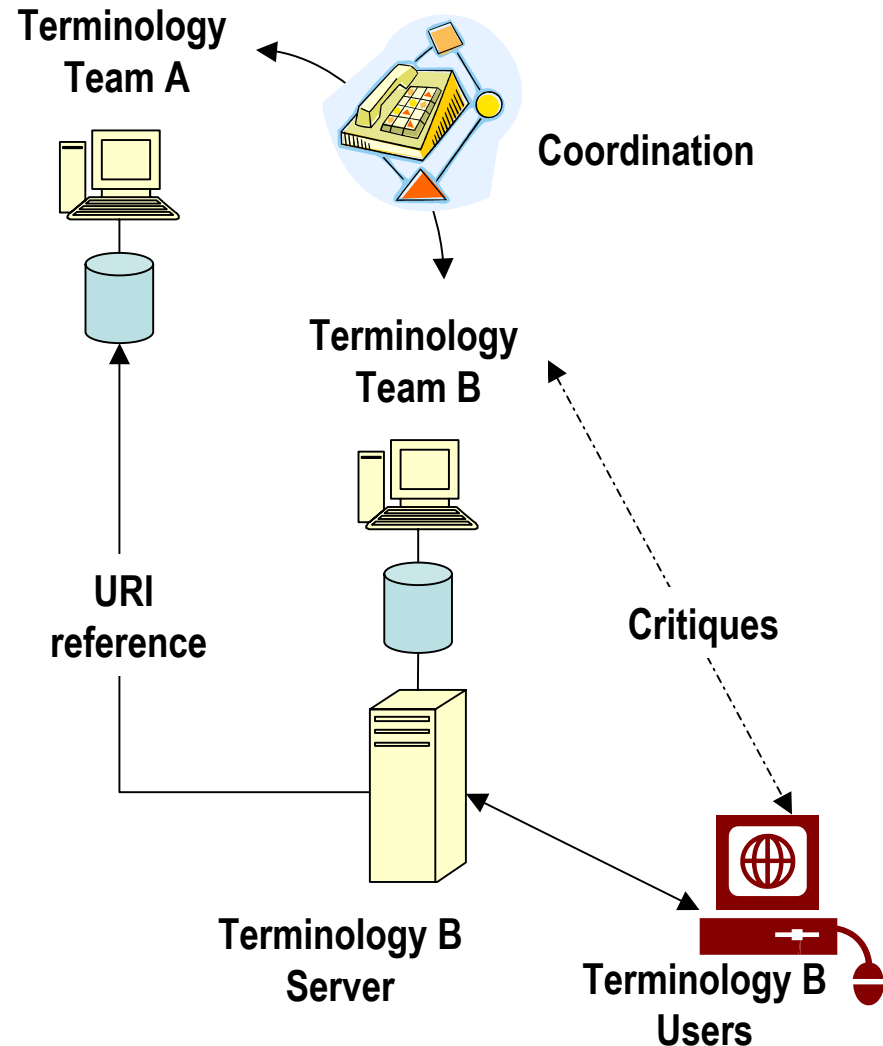- Some approaches

Centralized

- **Most common**
- **Advantages :**
  - **Common semantics**
  - **Granularity**
  - **Consistent quality**
  - **Fewer need to learn editing s/w**
- **Slow (usually)**
- **Expensive (always)**

Assignments/
Change Sets

Direction

Editorial Board

Central Site

Schema Synchronization

Critiques

Publication

Reviewers

Federation (loose)

- *Early* Semantic Web (e.g. NCI in MGED Ontology)

- **Disadvantages :**
  - **Nothing is guaranteed:**
    - **Common semantics**
    - **Granularity**
    - **Consistent quality**
  - **Assumes URI capable terminology servers**

- **Advantages :**
  - **Quick (usually)**
  - **Leverages what's there**
    - **Available terminologies**
    - **Subject matter expertise**

Terminology Team A

Coordination

Terminology Team B

URI reference

Critiques

Terminology B Server

Terminology B Users

Federation (formal)

- Distributed production
  - Focus on terminology for a subject matter domain
  - May offer terminology stand-alone
- Centralized integration
  - Integrates terminology
  - Focus point for users
    - Support, training
    - Publication
      - » Central Schema –or-
      - » RDF & Semantic Web
  - Assigns domains to hubs
  - Defines semantic model, style, granularity
- Disadvantages :
  - Untried
- Advantages :
  - May make federated terminologies practical
  - Hubs tied to community
    - Different rates of change
    - Short user-editor loop
  - May be coming thing
    - OBO, SOFG

Terminology Teams

A

B

C

Assigned Area

Central Authority

Updates for Master Schema

Central Site

Critiques

Critiques

Terminology C Server

Users of Terminology C

Users of Integrated Terminology